norden

Nordic Innovation Centre

POSITION PAPER

Position paper
Approved 2002-11

UNCERTAINTY - TO A CERTAIN LEVEL

NORDTEST
a Nordic Innovation Centre brand

Nordic Innovation Centre
Holbergsgate 1, 0166 OSLO
Telephone +47 23 35 45 40
Fax +47 23 35 45 45
info@nordicinnovation.net
www.nordicinnovation.net

ISSN:
Project:

1. Introduction

It is often stated that a value from a measurement has no meaning if the uncertainty is not known. This is a truism, of course. It has, however, in the view of the author, led to too far-reaching consequences in quality assurance systems and in accreditation requirements regarding analysis, demonstration and documentation of uncertainty in results from tests, analyses and measurements.

The present situation is that a laboratory often has to demonstrate in great detail the traceability and a budget of uncertainty not only for measurements and calibrations, but also for testing situations where these concepts are not fully relevant and have to be interpreted with care, making use of a sound understanding of the technical background of the test.

The standard requirement is to use the GUM [1] or similar codes, [2], [3], for the assessment of the uncertainty. In such codes it is presumed that all relevant factors influencing the uncertainty are known, that their magnitude can be assessed rather accurately, and that the functional relationship between each factor of uncertainty and the uncertainty of the end result is known. In many situations of testing and analysis, which are of great practical importance, these presumptions are not fulfilled. In others, the consequences of applying the code are very expensive and do not give added value for the client of the laboratory.

The intention with this position paper is to demonstrate a number of very common situations of testing and analysis where the uncertainty can be assessed efficiently and on a relevant level for the use of the client by a number of methods complementing [1] and similar codes. Such methods have indeed been taken into use, and to an increasing extent, but they should still be further emphasised and acknowledged, and employed more systematically.

2. Validation and traceability

In simple words, the validation of a test method means that it is demonstrated a) that the result of the test is relevant for answering the question asked (e.g. that a car crash test is a relevant simulation of a real crash), and b) that the uncertainty of the result is known and sufficiently small for the intended purpose. In this position paper, only the uncertainty part is discussed, with special attention to the various ways of assessing uncertainty which are useful in different situations.

Traceability in a measurement means that the relationship to the true value can be traced through a chain of comparative measurements, calibrations, from the measuring instrument used back to the realisation of the international standard, which is taken to represent a true value. One example is the measurement of voltage by a volt meter. This volt meter should be calibrated in advance by a chain of better and better volt meters, back to the national realisation of voltage through a Josephson reference. In each step, from the national standard and down, the uncertainty increases, since the use of simpler industrial instruments implies larger sources of uncertainty. Still, the end result should be that the voltage measured should not differ from the true value by more than the stated interval of uncertainty with a certain, stated probability.

The requirement that a measuring instrument is traceably calibrated to a stated uncertainty can be fulfilled in several ways, depending on what is required from the links in the traceablity chain, or, in other words, the accepted final level of unceratainty.

In what follows, the idealised descriptions of validation and traceability above are not questioned as such. It is demonstrated, however, that many situations exist which can be treated with good practical results by simpler procedures than those recommended or required at present.

3. Measurements and tests
It is important to distinguish between measurements on one hand and tests on the other. Much of the thinking behind the statistical approach in [1] is related to metrology and measurements, where the influencing factors are more clear-cut and where comparisons and repetitive measurements can be performed more easily than in testing. In the testing situation, there are often influencing factors which are only known very vaguely and the test result is a function of other factors than a number of measurements only.
Therefore the suggested procedures are presented in terms of measurements and tests separately. It must be noted then that measurements included in a testing procedure should, of course, follow what is recommended for those types of measurements.

4. A matrix of recommended procedures for various types of tests and measurements
In Figure 1 a number of situations of measurements and tests, and a number of suggested approaches for assessing uncertainty, is shown. The idea is to demonstrate a systematic approach for relating the type of test or analysis to a method for assessing uncertainty which is the most optimal one, sometimes with one or two options.
The first three cases concern measurements. Case a) is typically the situation where a step in a chain of traceability is realised, e.g. by calibration of the measurement standard of an accredited laboratory against a national standard. It could also be the situation of a precision measurement in a measurement laboratory. Here, the factors of uncertainty should be reasonably well known and the personnel should have the knowledge necessary to carry out the analysis in accordance with [1].
Case b) could be when a new measurement standard is to be introduced and a number of laboratories are involved in the establishment of the capability of the new standard. Depending on the state of development of the measurement standard either [1] or a profound statistical analysis could be utilised, or perhaps even both these approaches.

	Application of GUM [1] or similar code	Global inter-comparison, ISO/IEC-Guide 43 [4]	Stated un-certainty for instrument or method	In-house validation, fit for purpose	Exaggera-ted value, but fit for purpose	Statement that there is a "grey zone"
Calibration or single measurement						
a) High requirement on uncertainty, all relevant factors known	Preferred					
b) High requirement on uncertainty, all relevant factors not known	Optional	Preferred		Optional		
c) Low requirement on uncertainty (routine			Preferred		Optional, useful	

measurement)						
Standardised test or analysis with several measurements and influencing factors						
d) Uncertainty stated in standard			Preferred			
e) All relevant factors under control	Preferred			Optional, useful		
f) All relevant factors not under control		Optional		Preferred	Optional	Optional
Non-standardised test or analysis with several measurements and influencing factors						
g) Frequently used, all relevant factors controlled	Preferred	Optional		Optional		
h) Frequently used, all relevant factors nor under control		Optional		Preferred	Optional	
i) New or infrequently used, all relevant factors controlled	Optional			Preferred		
j) New or infrequently used, all relevant factors not under control				Preferred	Optional	Optional

Figure 1. Different situations of tests and measurements, and ways of assessing their uncertainty.

The case could lead to international/global interlaboratory work. A useful tool in these stages of development is Failure Mode and Effect Analysis (FMEA).
A very common situation in testing laboratories is represented by case c). This is when a measurement is made as a part of a test, as that of a length, a force, a weight or a time span.

Then it is often enough to use the standard value for uncertainty limit, e.g. a calibrated 0.5% load cell is just stated to have an uncertainty of 0.5%.
A possible point of conflict here is what can be considered as a "calibration" in many simple but important situations, and what is a necessary calibration interval. Generally it is recommended to adhere to the international recommendations, as ISO 10012 [5], but with common sense. Chains of in-house calibrations can also be recommended.
In some cases, as in c) in Figure 1, it is practical that the uncertainty estimates are stated as consciously exaggerated, so that their validity is self-evident. Normally the resulting interval of uncertainty is still quite sufficient. As typical examples of this situation may be taken the following two. First, the length markings of a steel ruler purchased from the iron-monger may be taken to be traceably calibrated for ever with an uncertainty of + 2 mm as long as no damage is visible. The original purchase control can be made by comparing with a similar ruler already existing in the laboratory. Second, the uncertainty of a normal wrist-watch can be taken to be less than 5 minutes in 24 hours, or less than 1 second in 5 minutes, if the bearer can use the watch daily for catching the bus and keeping appointments. One must, of course, be careful not to use many such exaggerated values in a formula combining them into a total uncertainty estimate. Such a procedure might give unnecessarily large estimates.

A majority of tests are standardised, i.e. their performance is described in a standard so that it shall be possible to carry them out in a reproducible way. This aim is not always achieved, since the description leaves room for interpretations, particularly in testing of the performance of products. Chemical analysis it is often described in terms of a methodology, rather than as a method, and then there is also ample room for deviations within the framework of such a methodology.
Still, in those cases, d) in the matrix of Figure 1, where the production of the standard has comprised a validation, and a standard value of uncertainty is given, this shall of course be used. This situation is desirable particularly in harmonised standards relating to EU directives. This is not yet so, and it is recommended that considerable work is devoted to this task, e.g. in the successor of the SMT program in the fifth framework program of R&D. This task is considered to be at least as important as the initiation of extensive programs for proficiency tests.

The case e) is typical for a material test, as e.g. the determination of fracture strength or fracture toughness of a metal. Here the testing situation normally is well-controlled and the measurements, length and force, can be performed with well-established uncertainties. The functional relationship between the measurements and the resulting test parameter is also known. The resulting uncertainty of the test parameter is also in general small enough in relationship to the normal variation in the material property and hence to the requirements for accuracy of the test result.

In the very common case of product testing the situation is much more complicated, and this is the typical situation covered by case f) in Figure 1. One may only mention the dubiously standardised area of toy testing, EN 71 [6]. Some very important areas are fire testing and EMC testing. Typical features here are that the property is a go/no go property of a complicated product (a functional disturbance is initiated or not in a computer by a certain electro-magnetic field strength etc.), and that the testing equipment (the furnace or the anechoic chamber) cannot be exactly the same in all laboratories. It should also be noted that test specimens are to some extent individuals and have differing properties, in the testing situation. Here a number of solutions have to be used depending on the requirements of the client or the authority. In some cases a value of the uncertainty which is easily seen to be larger than the factual one can be used, and in some cases it may be agreed that there is "grey zone" in which the uncertainty may lie, differing between laboratories. Such a "grey zone" may be useful in cases where there is a considerable gap between the limiting value for approval and the real value associated to the risk, as in some cases of environmental analysis.

It is recommended that groups of laboratories co-operate in order to find relevant measures of uncertainty in intercomparative studies. One such example is the IMEP programme in chemical analysis run by the IRMM in Geel (one of the Joint Research Centres of the EU). Experiences from this programme so far indicate that the uncertainty involved in even a rather clear-cut chemical analysis cannot be regularly produced by each single laboratory. A reliable value has to be determined, as a reference value, through extensive intercomparisons and statistical analysis. Otherwise, error components tend to be omitted or under-estimated.

It is worth noticing in this context that ISO 5725 [7] should be used for statistical evaluation of e g repeatability and reproducibility as soon as there is enough experimental data available from inter-comparisons or in-house investigations. It is also noted that as well intercomparisons as proficiency tests are useful general tools to establish realistic values of uncertainty. One must, however bear in mind the planning and aim of the investigation so that it is clear whether the laboratories, the test method or the group of test specimens are the subject for assessment. Often several aims are mixed up.

Finally, there is a large number of tests which are performed according to so called in-house methods. The laboratory puts together a testing programme to satisfy the requirements by clients or authorities, sometimes with parts which are common practise. Examples may be certain types of chemical analysis, or the implementation of general codes, such as the IEEE code for environmental and seismic qualification of electronic equipment. In the latter case the realisation and analysis of seismic disturbances is left to the laboratory within certain frames of tolerances in terms of a power spectrum.

If the test is performed frequently and experience is gained continually it may be worthwhile to collect the experience systematically, perhaps in co-operation with some other laboratories, cases g) and h) in Figure 1. Again, case g) is typical for physical material testing, and case h) for product testing and qualitative testing. With regard to the last-mentioned item, the concept of uncertainty has to be interpreted in a special fashion, meaning something like the risk for approving a "bad product" or disapproving a "good product". This can be handled by so called OC-curves, and this is a special subject with a well developed theory which can and should be employed. It is important to stress that a single laboratory should not be required to produce a full uncertainty budget for the type h) of tests. This is sometimes the case today and it causes much concern.

This is even more important to stress in the cases i) and j) in Figure 1, relating to new or infrequently used test methods. Such methods are often important for the technical development and progress in industry and society. They should be allowed and used, and not hampered by rigid requirements. Common sense should be used to weigh risks against benefits in the introductory stage. Standardisation is often a very slow process, meaning that standardised methods tend to be somewhat old-fashioned.

The case i) might seem too optimistic. One example, however, may be the evaluation of fracture toughness of a new composite plastic with a method, which is originally intended only for metals and which has a basically different set of constitutive material properties, as stiffness and visco-elasticity. Then, with a good solid mechanics knowledge in the laboratory, the uncertainty can be assessed and demonstrated properly.

The cases h) and j) are of course the most common ones. They are of great importance and they cause great concern. Particularly in the present development of energy technology and environmental protection there are many new types of tests and analyses which should be introduced and used, not least for indicative purposes. At the same time there are immense economical and health interests at stake. Again it is important to accept alternative ways of demonstrating uncertainty which are fit for purpose, and to initiate inter-comparative studies.
There should be an inclination in the future to rely more on laboratory competence and knowledge, and to promote quality management of such factors, and to rely less on the schematic calculation of uncertainty budgets in cases where this methodology does not lead to trustworthy results or become very expensive.

5. Summary and conclusions
It has been shown by a number of examples that attempts to demonstrate uncertainty through the presently recommended codes, as [1], are less feasible in many important applications of laboratory work. They also tend to become expensive and they do not in general give added value to the customers.

There are types of tests where an assessment of uncertainty is meaningless except as a very coarse estimate. There are so called go/no go tests which have to be terated in a special fashion. There are also tests where the natural dispersion in properties of test specimens is considerably larger than the uncertainty of the test method. In all these cases there is a need for alternatives.

A scheme of different testing and measurement situations has been suggested together with a number of principal, alternative ways to evaluate and state or indicate uncertainty. These ways are thought to be suitable to employ for the needs of clients. They are also thought to be necessary for enhancement of the technical and economical development of the laboratory industry, and for efficient support to industry and society from laboratories.

These alternatives require a partly new thinking in standardisation and accreditation bodies, and more focus in the laboratories on management of competence. They also require that much more efforts and resources are put into scientific work and into inter-comparisons to evaluate the properties of methods for testing and analysis.

6. References

[1] GUM: Guide to the expression of uncertainty in measurement, BIPM/IEC/IFCC/ISO/ OIML/IUPAC/IUPAP, 1995.
[2] EAL-R2: "Expression of the Uncertainty of Measurement in Calibration", EAL, 1997.
[3] "Quantifying Uncertainty in Analytical Meaurement", EURACHEM, 1996.
[4] ISO/IEC Guide 43 - (1,2): 1997, Proficiency testing by interlaboratory comparisons, 1997.
[5] ISO 10012 - 2: 1997 (E). Quality assurance for measuring equipment - Part 2: Guidelines for control of measurement processes, 1997.
[6] EN 71 - (1,2). 1993, Safety of toys, 1993.
[7] ISO 5725 - (1,6): 1994, Accuracy (trueness and precision) of measurement methods and results, 1994.

POSITION PAPER 005
Approved 1998-06

Nordic Innovation Centre

Return address:
Nordic Innovation Centre,
Holbergs gate 1,
NO-0166 Oslo, Norway

NORDTEST

NORDTEST is a Nordic Innovation Centre brand offering competence and expertice in the field of harmonizing of norms and methods,a large Nordic net-work of experts, more than 650 recommended Nordic testing methods and 550 published technical reports.

www.nordicinnovation.net

Nordic Innovation Centre

The Nordic Innovation Centre initiates and finances activities that enhance innovation collaboration and develop and maintain a smoothly functioning market in the Nordic region.

The Centre works primarily with small and medium-sized companies (SMEs) in the Nordic countries. Other important partners are those most closely involved with innovation and market surveillance, such as industrial organisations and interest groups, research institutions and public authorities.

The Nordic Innovation Centre is an institution under the Nordic Council of Ministers. Its secretariat is in Oslo.

For more information: www.nordicinnovation.net